International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-14 Issue-8, July 2025

Protocol for Vision-Based Tracking and Proportional Control in Quadcopter Follow-Me Applications

Shayak Bose, Chethana G

***Abstract:** This paper presents the design and implementation of a real-time "Follow Me" protocol for a quadcopter using vision-based tracking and proportional control. The system enables a drone to autonomously follow a moving human subject using bounding box detections from an onboard AI-based object detection camera stream. The design extracts the width of the bounding box surrounding the target and uses it as a reference for distance. A proportional control algorithm maps the deviation of the observed width from a predefined ideal width into a corresponding pitch velocity, which is then converted to PWM signals to drive the drone. The control logic ensures that the drone maintains an optimal distance from the subject by dynamically adjusting its forward and backwards movement. Experimental results demonstrate a linear and monotonic relationship between the bounding box width and the drone's pitch signal, validating the accuracy and responsiveness of the tracking system. The proposed system operates robustly in real-time and can be integrated into lightweight UAV platforms without requiring GPS or external localisation systems.*

***Index Terms:** Quadcopter, Follow-Me Protocol, Visual Tracking, Bounding Box Width, Proportional Control, PITCH PWM, UAV Control, Real-Time Drone System*

Abbreviations:
UAVs: Uncrewed Aerial Vehicles

I. INTRODUCTION

Uncrewed Aerial Vehicles (UAVs), particularly quadcopters, have gained significant popularity in various applications, including surveillance, delivery systems, agriculture, and human interaction tasks. An emerging capability in this domain is the" Follow-Me" protocol, where a UAV autonomously tracks and follows a human subject based on visual and positional cues. This functionality finds application in autonomous filming, rescue operations, surveillance, and personal assistance systems. Traditional follow-me implementations often rely on GPS trackers or external localization systems. However, these solutions face limitations in indoor or GPS-denied environments.

Manuscript received on 29 May 2025 | First Revised Manuscript received on 06 June 2025 | Second Revised Manuscript received on 20 June 2025 | Manuscript Accepted on 15 July 2025 | Manuscript published on 30 July 2025.
*Correspondence Author(s)
Shayak Bose*, Department of Electronics and Communication Engineering, RV College of Engineering, Bengaluru (Karnataka), India. Email ID: shayakbose.ec21@rvce.edu.in, ORCID ID: 0009-0000-3216-2811
Dr. Chethana G, Department of Electronics and Communication Engineering, RV College of Engineering, Bengaluru (Karnataka), India. Email ID: chethanag@rvce.edu.in

environments. To address such challenges, recent research has shifted toward vision-based tracking methods that employ onboard cameras. These systems utilise object detection and tracking techniques to identify and track humans in real-time, thereby eliminating the need for external positioning infrastructure.

Recent works have incorporated deep learning and transformer-based models to enhance the robustness of tracking. For example, Chen et al. [1] proposed a vision-based autonomous target-following system for quadcopter drones, demonstrating reliable performance in outdoor environments.

Similarly, Barisic et al. [2] developed a real-time detection and following system that operates effectively in dynamic scenarios. Advanced tracking algorithms have been further explored by researchers, with Pan et al. [3] introducing Fast-Tracker 2.0, which improves autonomy through active vision and human location regression techniques.

The integration of sophisticated visual tracking methodologies has led to enhanced performance metrics. Li et al. [4] developed AutoTrack, a high-performance visual tracking system that incorporates automatic spatio-temporal regularisation for UAV applications. Additionally, collective behaviour studies by Schilling et al. [5] have demonstrated vision-based drone flocking capabilities in outdoor environments, expanding the potential for multi-UAV coordination systems.

Several studies have focused on the practical implementation aspects of autonomous human following systems. Piquero et al. [6] presented a novel implementation using local context analysis. At the same time, advanced vision transformer approaches have been explored by Yao et al. [7] through their SGDViT framework for UAV tracking applications. Wu et al. [8] contributed to real-time aerial object localisation and tracking methodologies, demonstrating the feasibility of vision-based sensing systems.

Early implementations of vision-based tracking systems have established foundational approaches for UAVs following humans. Kim and Shim [9] developed a vision-based target tracking control system for quadrotors using tablet computer interfaces, demonstrating the integration of mobile computing platforms with aerial vehicles. Adekola et al. [10] further explored object tracking-based follow-me UAV systems, providing a comprehensive analysis of tracking algorithms for autonomous following applications.

Implementation studies have validated the practical feasibility of person tracking in drone systems.

Nordberg et al.

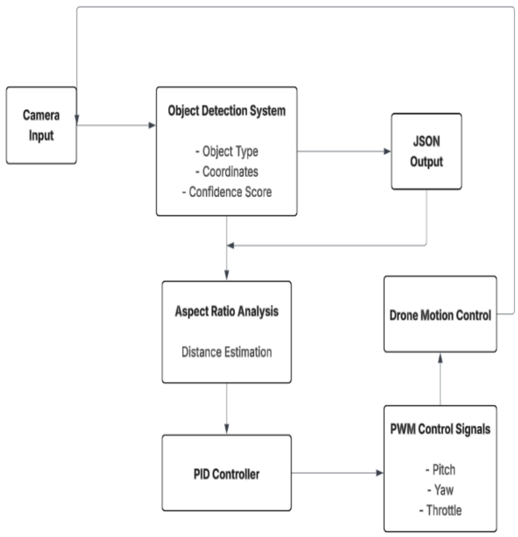
Conducted detailed implementation research on video-based person tracking [11], confirming real-world deployment capabilities. Additionally, Naseer et al. [12] introduced person-following and gesture-recognition systems for quadcopters, expanding the interaction modalities beyond simple following to include human-gesture interpretation. Building on this direction, this work presents a vision-based follow-me protocol that enables a quadcopter to follow.

A human subject in real time using only onboard video feeds. The system detects the target using a bounding box generated by a person detection model and uses the width of this box as a proxy for the distance between the drone and the subject. A proportional control strategy is applied to compute the pitch velocity, which is then mapped to a PWM signal for drone actuation. This ensures that the drone dynamically adjusts its forward or backwards motion to maintain a fixed separation from the subject.

Experimental validation demonstrates a consistent and approximately linear relationship between the bounding box width and the pitch signal sent to the UAV, confirming the reliability and responsiveness of the proposed control system. The method is lightweight, operates on embedded platforms, and is designed to function without any external localization aids, making it ideal for deployment in constrained and GPS-denied environments.

II. METHODOLOGY

In this section, the process, along with the necessary steps and equations, has been discussed to obtain the desired results. The system design follows a modular pipeline, beginning with image acquisition and concluding with control signal generation. Each module is designed to process input data sequentially, enabling real-time decision-making. Detailed explanations of the flow diagram, component responsibilities, and interconnections are provided to ensure clarity and reproducibility of the proposed Follow Me Protocol.



[Fig.1: Flowchart of Follow Me Protocol]

The proposed **Follow Me Protocol** is designed to enable the autonomous tracking of a human subject by a UAV (Unmanned Aerial Vehicle) using onboard vision systems and control algorithms. The flow of information and control signals in the system is represented in Figure 1 and elaborated as follows:

A. Camera Input

The system begins with continuous video input from a camera mounted on the UAV. This stream serves as the primary sensing modality, capturing the environment in real-time for further processing.

B. Object Detection System

The captured video frames are processed by a deep learning-based Object Detection System, which performs:

i. *Object Type Identification:* Determines if the object of interest (e.g., a person) is present.
ii. *Coordinate Extraction:* Provides bounding box coordinates (x, y, width, height) of the detected object.
iii. *Confidence Score:* Outputs a probability score indicating the reliability of detection.

C. JSON Output

The detection results are serialised into a JSON format that contains metadata, including object class, bounding box dimensions, and confidence score. This allows for lightweight, structured communication and logging.

D. Aspect Ratio Analysis

To estimate the distance between the UAV and the tracked person, the Aspect Ratio Analysis module evaluates the bounding box dimensions. Changes in the bounding box size are used to infer how far or close the subject is moving from the UAV (depth estimation).

E. PID Controller

The error between the desired and actual dis- tance (and position) is passed into a PID (Propor- tional–Integral–Derivative) Controller, which computes corrective action. This ensures smooth and stable following behaviour by minimising overshoot and oscillations.

F. PWM Control Signals

Based on the PID output, PWM (Pulse Width Modulation) signals are generated to adjust drone flight parameters:

i. *Pitch* (forward/backwards tilt)
ii. *Yaw* (rotation to left/right)
iii. *Throttle* (altitude control)

These signals provide fine-grained control of the drone's movement.

G. Drone Motion Control

This module receives the PWM signals and interfaces directly with the UAV's motors to adjust its motion. It also ensures flight stability and obstacle avoidance (if applicable).

H. Feedback Loop

There is a continuous feedback loop:

i. The updated position from Drone Motion Control is fed back into the Object Detection System to reacquire the target.

ii. The JSON Output can optionally provide external interfaces for diagnostics or remote monitoring.

The protocol also uses a few equations to calculate the appropriate PWM to maintain a distance between the" person" and the Quadcopter:

III. RESULTS AND ANALYSIS

The results obtained using the proposed control equations.

$$best = \max(targets, key = \text{lambda } d: d.get("box", [0,0,0,0])[2]) \quad \dots \quad (1)$$

$$x, y, h, w = best.get("box", [0,0,0,0]) \quad \dots \quad (2)$$

$$C_x = x + \frac{w}{2} \quad \dots \quad (3)$$

Demonstrate a high degree of consistency and accuracy in tracking performance. Notably, the relationship between the PITCH PWM signal and the width and height of the detected bounding box exhibits an approximately linear trend, as expected from the proportional control logic. This consistency

From equations (1), (2), (3), the width is extracted from the JSON Output over height due to better **Field of View** reliability, and then its centre is calculated.

$$err_x = \frac{c_x - \frac{\text{IMG}w}{2}}{\frac{\text{IMG}_W}{2}} \quad \dots \quad (4)$$

$$err_{dist} = \frac{DES\ BOX_W}{DES\_BOX_W} w - w \quad \dots \quad (5)$$

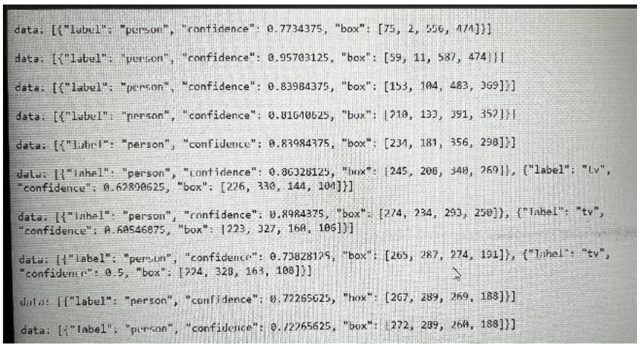
Validates the effectiveness of the visual feedback–based navi- gation strategy implemented in the system.

Next, using equations (4) and (5), the error distances based on the reference Bounding Box Width are measured.

$$v_y = -K_p^{yaw} \cdot err_x \quad \dots \quad (6)$$

$$v_x = K_p^{yaw} \cdot err_{\text{dist}} \quad \dots \quad (7)$$

Finally, the lateral and forward velocity components, v_y and



data: [{"label": "person", "confidence": 0.7734375, "box": [75, 2, 556, 474]}]
data: [{"label": "person", "confidence": 0.95703125, "box": [59, 11, 587, 474]}]
data: [{"label": "person", "confidence": 0.83984375, "box": [153, 104, 483, 369]}]
data: [{"label": "person", "confidence": 0.81640625, "box": [210, 133, 391, 352]}]
data: [{"label": "person", "confidence": 0.83984375, "box": [234, 181, 356, 298]}]
data: [{"label": "person", "confidence": 0.86328125, "box": [245, 208, 340, 269]}, {"label": "tv", "confidence": 0.62890625, "box": [226, 330, 144, 104]}]
data: [{"label": "person", "confidence": 0.8984375, "box": [274, 234, 293, 250]}, {"label": "tv", "confidence": 0.60546875, "box": [223, 327, 160, 106]}]
data: [{"label": "person", "confidence": 0.73828125, "box": [265, 287, 274, 191]}, {"label": "tv", "confidence": 0.5, "box": [224, 328, 163, 108]}]
data: [{"label": "person", "confidence": 0.72265625, "box": [267, 289, 269, 188]}]
data: [{"label": "person", "confidence": 0.72265625, "box": [272, 289, 260, 188]}]

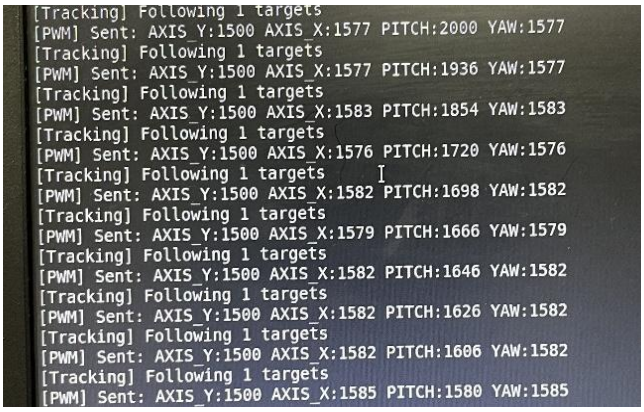
[Fig.2: Real-Time Person Detection Output with Bounding Box Parameters]

Vx are computed using equations (6) and (7), respectively. The forward velocity, v_x, is **linearly dependent on the deviation of the bounding box width from a desired setpoint**. Specifically, it is proportional to the term (DESBOXW-w), where *w* is the detected bounding box width. Although the controller internally computes forward velocity as $v_x \propto$ (DESBOXW-w), this value is negated before being converted to a PWM command (Equations (8) and (9)). Hence, the actual PITCH PWM signal becomes **directly proportional** to the bounding box width *w*, ensuring that the Quadcopter moves backwards as the person gets closer (i.e., when *w* increases) and forward when the person is farther (i.e., when *w* decreases).

$$velocity\_to\_pwm\left(\max\left(-1.0, \min\left(1.0, vy\right)\right)\right) \quad \dots \quad (8)$$

$$velocity\_to\_pwm\left(\max\left(-1.0, \min\left(1.0, -vx\right)\right)\right) \quad \dots \quad (9)$$

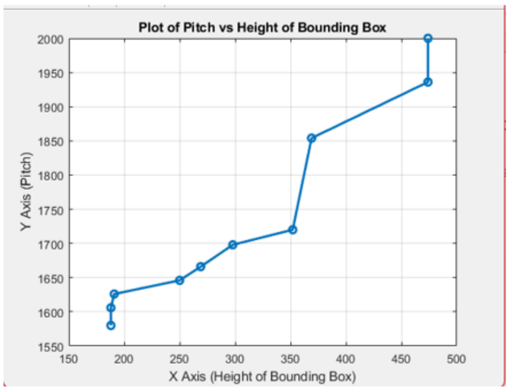
Equations (8) and (9) also serve to normalize and scale the computed velocity values into PWM-compatible signals. The clamping operation within the velocity_to_pwm function ensures that extreme or noisy velocity inputs do not exceed the expected control bounds. As a result, the generated PWM signals remain within a safe and linear operating range of 1000 to 2000 microseconds, centred at 1500. This translation from visual feedback to motor command allows for smooth and bounded control, enabling the Quadcopter to respond proportionally and reliably to real-time changes in the subject's position and distance.



```
[Tracking] Following 1 targets
[PWM] Sent: AXIS_Y:1500 AXIS_X:1577 PITCH:2000 YAW:1577
[Tracking] Following 1 targets
[PWM] Sent: AXIS_Y:1500 AXIS_X:1577 PITCH:1936 YAW:1577
[Tracking] Following 1 targets
[PWM] Sent: AXIS_Y:1500 AXIS_X:1583 PITCH:1854 YAW:1583
[Tracking] Following 1 targets
[PWM] Sent: AXIS_Y:1500 AXIS_X:1576 PITCH:1720 YAW:1576
[Tracking] Following 1 targets
[PWM] Sent: AXIS_Y:1500 AXIS_X:1582 PITCH:1698 YAW:1582
[Tracking] Following 1 targets
[PWM] Sent: AXIS_Y:1500 AXIS_X:1579 PITCH:1666 YAW:1579
[Tracking] Following 1 targets
[PWM] Sent: AXIS_Y:1500 AXIS_X:1582 PITCH:1646 YAW:1582
[Tracking] Following 1 targets
[PWM] Sent: AXIS_Y:1500 AXIS_X:1582 PITCH:1626 YAW:1582
[Tracking] Following 1 targets
[PWM] Sent: AXIS_Y:1500 AXIS_X:1582 PITCH:1606 YAW:1582
[Tracking] Following 1 targets
[PWM] Sent: AXIS_Y:1500 AXIS_X:1585 PITCH:1580 YAW:1585
```

[Fig.3: Real-Time PWM Outputs for Detected Targets]

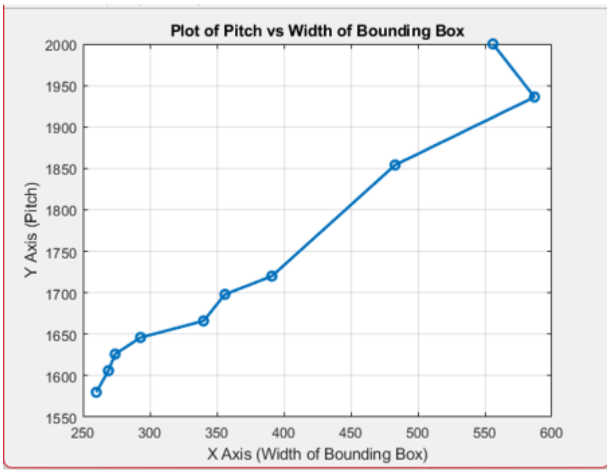
Figures 2 and 3 illustrate the real-time JSON detection outputs and their corresponding PITCH PWM control signals during a live quadcopter flight. These outputs reflect how the system dynamically adapts motor commands in response to the subject's relative position and distance.



[Fig.4: Plot of PITCH PWM vs. Height of Bounding Box]

Retrieval Number: 100.1/ijitee.G110714070625
DOI: 10.35940/ijitee.G1107.14080725
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)



[Fig.5: Plot of PITCH PWM vs. Width of Bounding Box]

Figures 4 and 5 present the observed variation in PITCH PWM values concerning the height and width of the bounding box, respectively. Both plots show a strong linear correlation, confirming that the quadcopter adjusts its forward or backwards motion proportionally based on the perceived size of the subject. This behavior is consistent with the control equations discussed earlier and highlights the reliability of the follow-me protocol in translating visual cues into effective actuation commands. The linearity in both dimensions further supports the robustness of the implemented vision-based distance estimation and control methodology.

IV. CHALLENGES AND LIMITATIONS

While the proposed Follow-Me protocol demonstrates effective real-time person tracking and drone control using visual feedback, several challenges and limitations were identified during development and testing.

A key limitation arises from the system's reliance on the bounding box width as the primary proxy for estimating the distance between the drone and the subject. This approach, though effective in most scenarios, can be influenced by changes in body orientation, occlusion, or non-frontal poses, which may distort the apparent width and lead to inaccurate results.

Distance estimation. Although the plot of PITCH versus binding box height also exhibited a consistent trend, it was not used in the control loop due to practical stability and simplicity considerations.

Lighting variability was another significant challenge. Detection performance deteriorated in low-light or high-glare conditions, occasionally causing the UAV to issue neutral or delayed commands. Integrating infrared sensing or adaptive exposure control could enhance robustness under varying illumination conditions.

The system currently tracks the largest detected bounding box, assuming it corresponds to the intended target. This assumption may not hold in crowded or multi-subject environments, potentially causing tracking ambiguity. Incorporating identity tracking or gesture-based target selection could mitigate such issues.

From a control standpoint, the use of a proportional-only control strategy, while computationally efficient, does not compensate for UAV inertia or dynamic overshoot. This can result in sluggish or oscillatory responses during rapid target motion. Implementing a full PID or adaptive controller may improve stability and precision.

Lastly, the current system has been validated in controlled indoor settings. Outdoor environments present additional challenges such as wind, variable lighting, and GPS drift, all of which require further adaptation and testing for robust field deployment.

V. CONCLUSION AND FUTURE WORK

In this work, a vision-based Follow-Me protocol for quadcopters is proposed and implemented, utilising real-time person detection and bounding box analysis. The system utilizes the width of the detected bounding box as a proxy for the subject's distance and computes corresponding control signals using a proportional control strategy. The resulting PWM values are mapped accurately to drone actuation commands, allowing the UAV to maintain an appropriate distance from the target.

Experimental results, including real-time output logs and plotted response graphs, demonstrate a strong linear correlation between the PITCH PWM signal and the bounding box dimensions, validating the robustness and effectiveness of the approach. The simplicity of the control logic, combined with the system's ability to operate without external localisation systems, makes it suitable for lightweight embedded deployment.

For future work, several enhancements are envisioned. Integrating a PID or adaptive control scheme could provide smoother and more stable flight behaviour, especially in dynamic environments. Incorporating additional cues such as optical flow, depth estimation, or re-identification mechanisms could improve tracking performance in cluttered scenes or multi-subject scenarios. Moreover, extending the system to operate reliably under varying outdoor conditions, including variable lighting and wind disturbances, is an essential step toward achieving deployment-ready autonomy.

DECLARATION STATEMENT

After aggregating input from all authors, I must verify the accuracy of the following information as the article's author.

- **Conflicts of Interest/ Competing Interests:** Based on my understanding, this article has no conflicts of interest.
- **Funding Support:** This article has not been funded by any organizations or agencies. This independence ensures that the research is conducted with objectivity and without any external influence.
- **Ethical Approval and Consent to Participate:** The content of this article does not necessitate ethical approval or consent to participate with supporting documentation.
- **Data Access Statement and Material Availability:** The adequate resources of this article are publicly accessible.
- **Authors' Contributions:** Each author has individually contributed to the article. SHAYAK BOSE is the primary author of the article and is involved in the design: experiments, research, analysis &

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

writing of the article. Dr CHETHANA G has assisted as a guide from RVCE and has helped with proofreading and providing invaluable suggestions throughout the process.

REFERENCES

1. W.-C. Chen, C.-L. Lin, Y.-Y. Chen, and H.-H. Cheng, "Quadcopter Drone for Vision-Based Autonomous Target Following," *Aerospace*, vol. 10, no. 1, p. 82, Jan. 2023, DOI: http://doi.org//10.3390/aerospace10010082
2. A. Barisic, M. Car, and S. Bogdan, "Vision-Based System for a Real-Time Detection and Following of UAV," *arXiv preprint arXiv:2205.00083*, Apr. 2022. [Online]. Available: https://arxiv.org/abs/2205.00083
3. N. Pan, R. Zhang, T. Yang, C. Xu, and F. Gao, "Fast-Tracker 2.0: Improving Autonomy of Aerial Tracking with Active Vision and Human Location Regression," *arXiv preprint arXiv:2103.06522*, Mar. 2021. [Online]. Available: https://arxiv.org/abs/2103.06522
4. Y. Li, C. Fu, F. Ding, Z. Huang, and G. Lu, "AutoTrack: Towards High- Performance Visual Tracking for UAV with Automatic Spatio-Temporal Regularization," *arXiv preprint arXiv:2003.12949*, Mar. 2020. [Online]. Available: https://arxiv.org/abs/2003.12949
5. F. Schilling, F. Schiano, and D. Floreano, "Vision-Based Drone Flocking in Outdoor Environments," *arXiv preprint arXiv:2012.01245*, Dec. 2020. [Online]. Available: https://arxiv.org/abs/2012.01245
6. J. Piquero, E. Sybingco, A. Chua, M. Say, C. Crespo, R. Rivera, M. A. Roque, and L. Ambata, "A Novel Implementation of an Autonomous Human Following Drone using Local Context," *Int. J. Autom. Smart Technol.*, vol. 11, no. 1, pp. 1–7, 2021, DOI: http://doi.org//10.5875/ausmt.v11i1.2147
7. L. Yao, C. Fu, S. Li, G. Zheng, and J. Ye, "SGDViT: Saliency-Guided Dynamic Vision Transformer for UAV Tracking," *arXiv preprint arXiv:2303.04378*, Mar. 2023. [Online]. Available: https://arxiv.org/abs/2303.04378
8. Y. Wu, Y. Sui, and G. Wang, "Vision-based Real-Time Aerial Object Localization and Tracking for UAV Sensing System," *arXiv preprint arXiv:1703.06527*, Mar. 2017. [Online]. Available: https://arxiv.org/abs/1703.06527
9. J. Kim and D. H. Shim, "A Vision-Based Target Tracking Control System of a Quadrotor by Using a Tablet Computer," in *Proc. Int. Conf. Unmanned Aircraft Systems (ICUAS)*, 2013, pp. 1165–1170. [Online]. Available: DOI: https://doi.org/10.1109/ICUAS.2013.6564808
10. O. D. Adekola et al., "Object Tracking-Based 'Follow-Me' Unmanned Aerial Vehicle (UAV) System," Comput. Syst. Sci. Eng., vol. 41, no. 3, pp. 875–890, 2022. DOI: https://doi.org/10.32604/csse.2022.021029
11. E. Nordberg et al., "Implementation of Video-based Person Tracking in a Drone System," Independent thesis, 2021. [Online]. Available: https://www.diva-portal.org/smash/get/diva2:1563281/FULLTEXT01.pdf
12. T. Naseer, J. Sturm and D. Cremers, "FollowMe: Person following and gesture recognition with a quadrocopter," 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, Tokyo, Japan, 2013, pp. 624-630, DOI: http://doi.org//10.1109/IROS.2013.6696416

AUTHOR'S PROFILE

Mr. Shayak Bose is currently pursuing a Bachelor of Engineering (B.E.) in Electronics and Communication Engineering at RV College of Engineering (RVCE), Bangalore, India. He has a keen interest in the Digital and Embedded Systems domain, with several projects undertaken in this field. Mr. Bose is currently an intern at Signaltron Systems, serving as an Embedded Software Design Engineer, where he is gaining valuable experience and knowledge in 5G technology and its associated protocols. During his internship, he has also worked on GNSS for drones, contributing to advancements in drone navigation systems and the integration of AI into drone cameras for various applications such as object detection and autonomous flight. He is focused on enhancing his skills in embedded programming, system design, and network protocols.

Dr. Chethana G holds a B.E. in Electronics and Communication Engineering, an M.Tech. in Digital Communication and Networking, and a Ph.D. in MIMO (Multiple Input Multiple Output) systems, all in the field of electronics and computer networks, from the University of Mysore (VTU), Karnataka, India. She is currently serving as an Assistant Professor in the Department of Electronics and Communication Engineering at RV College of Engineering (RVCE), Bengaluru, India. Dr. Chethana has guided 20 undergraduate (UG) and five postgraduate (PG) projects. She has contributed to 15 journal publications and 20 conference papers. Her research interests span Wireless Sensor Networks (WSN), Cryptography, and Artificial Intelligence/Machine Learning (AI/ML).